

Le contrôle de l'information macromoléculaire

Résumé L'information est partout, aussi bien au cœur de nos cellules que dans nos téléphones. Ainsi, le développement de nouvelles technologies permettant le stockage et le transfert d'information est devenu un enjeu scientifique majeur. Dans ce contexte, il a été récemment montré que l'information – qu'elle soit numérique ou d'autre nature – peut être stockée dans des macromolécules contenant des séquences contrôlées de monomères. Si ce principe est connu depuis longtemps en biologie, il est nouveau dans le domaine des polymères synthétiques. Dans cet article, les approches chimiques permettant d'écrire, de lire et de modifier l'information macromoléculaire sont présentées. Les applications technologiques découlant de ce nouveau domaine de recherche sont brièvement discutées. Des perspectives inédites, notamment dans le domaine de la vie artificielle sont également évoquées en fin d'article.

Mots-clés Chimie des polymères, polymères numériques, stockage d'information, séquençage, biologie artificielle.

Abstract On the control of macromolecular information

Information is everywhere. It is stored in the nucleus of our cells but also in the memories of our phones. Thus, the development of information storage and information transfer technologies has become a central topic in fundamental and applied science. In this context, it was recently proposed that information can be stored in sequence-controlled macromolecules. Although this principle is long-known in molecular biology, it has only recently been developed in synthetic polymer science. In this article, polymer chemistry approaches allowing to write, read, and edit macromolecular information are presented. Furthermore, potential technological applications of this new class of polymers are briefly discussed. Last but not least, the hypothetical but potential use of synthetic informational polymers for the development of artificial life systems is critically analyzed.

Keywords Polymer chemistry, digital polymers, information storage, sequencing, artificial biology.

Le XXI^e siècle est sans aucun doute le siècle de l'information. En effet, les grandes avancées scientifiques et technologiques de ce début de siècle sont toutes liées d'une manière ou d'une autre à la maîtrise de l'information. Par exemple, au cours de ces deux dernières décennies, l'émergence de nouvelles technologies numériques a révolutionné le monde du travail, des loisirs et de la communication [1]. Mais les progrès de ce siècle ne se limitent pas aux domaines de l'informatique et des télécommunications : des avancées majeures ont aussi été obtenues dans le domaine de la biologie et, en particulier, de la génomique. Ainsi, le XXI^e siècle a débuté avec le séquençage complet du génome humain [2-3], découverte qui a ouvert de nouveaux horizons pour la médecine et les biotechnologies, mais aussi dans d'autres champs disciplinaires tels que l'archéologie, l'histoire et la généalogie. Dans ce cas, l'information qui est traitée n'est pas numérique mais génétique, et à l'instar d'une séquence binaire stockée sur un disque dur, cette information biologique peut désormais être interprétée, copiée, créée et modifiée [4-5].

Cependant, le contrôle de l'information n'est pas l'apanage des informaticiens et des biologistes ; c'est aussi le travail du chimiste. En effet, pour contrôler l'information, il faut aussi contrôler la matière. Plus précisément, il faut maîtriser les matériaux et les phénomènes physiques qui permettent le stockage de l'information ainsi que son transfert et son interprétation. Chimiquement et physiquement parlant, ces supports d'information peuvent être de natures très diverses. Par exemple, les phénomènes mis en jeu dans une mémoire d'ordinateur, une mémoire cérébrale ou un noyau cellulaire sont radicalement différents. Dans le dernier cas, le message génétique est stocké sur une macromolécule, l'ADN, qui est en quelque sorte le disque dur de la vie. L'ADN est un polymère fabriqué à l'aide de quatre briques élémentaires (ou monomères) qui sont les nucléotides A, T, G et C. L'information

génétique y est stockée sous forme d'une séquence contrôlée de monomères [6]. En d'autres termes, l'ADN permet un stockage « macromoléculaire » de l'information.

Depuis quelques années, l'équipe que j'anime à l'Institut Charles Sadron de Strasbourg a contribué à démontrer que la notion d'information macromoléculaire n'est pas propre à l'ADN, ni même à la biologie en général. Dans notre perception scientifique du monde, l'ADN a un statut à part car il s'agit du support de la vie. Toutefois, sous le regard plus neutre et plus technique du polymériste, il est possible de « désacraliser » l'ADN et de ne le voir que comme un polymère comme un autre. Tous les polymères, qu'ils soient biologiques ou fabriqués par l'homme, sont des chaînes (le plus souvent linéaires) formées par l'association d'unités monomères. Ces briques moléculaires peuvent être toutes les mêmes (on parle alors d'homopolymère), ou légèrement différentes comme dans le cas de l'ADN (on parle alors de copolymère).

D'un point de vue purement théorique, absolument tous les copolymères semblent propices au stockage d'information, comme l'a souligné dès 1986 Richard Dawkins dans son livre *L'horloger aveugle* où il écrit : « *S'il y a deux sortes de petites molécules dans la chaîne, on peut les considérer respectivement comme 1 et 0, et une quantité quelconque d'information de n'importe quelle nature peut être immédiatement stockée pourvu que la chaîne soit assez longue* » [7]. En effet, une séquence ordonnée de monomères semble propice au stockage de n'importe quel type de message, qu'il soit écrit en langage génétique, informatique ou autre [8]. Cependant, si la possibilité d'un stockage macromoléculaire a été imaginée il y a assez longtemps, cette idée n'a été concrétisée que très récemment. La raison principale à cela est que les approches classiques de synthèse des polymères ne permettent pas de contrôler précisément les séquences de monomères, tout au moins pas avec un niveau de précision suffisant pour former

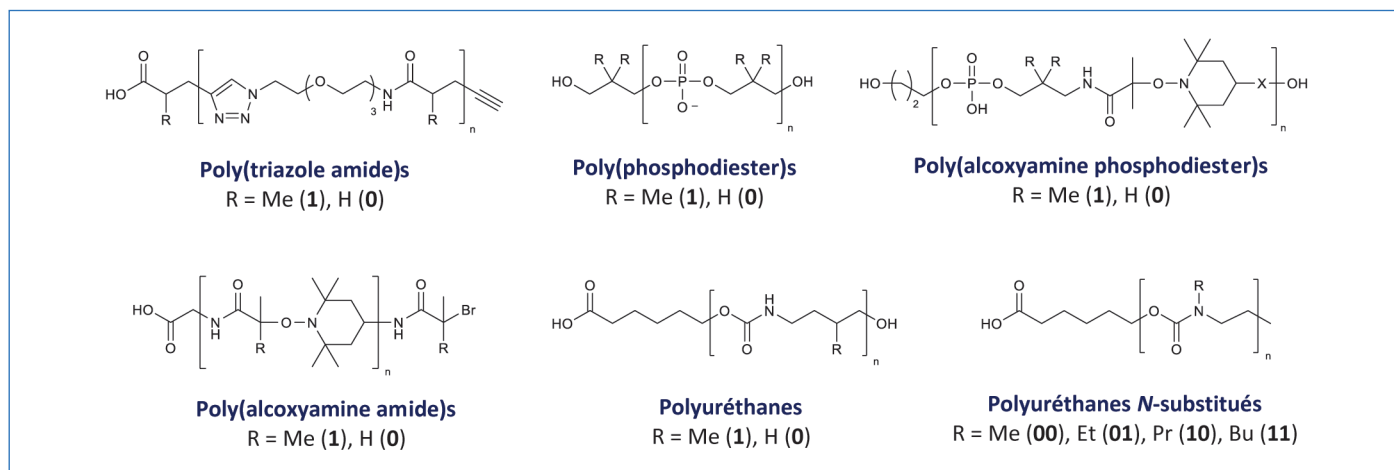


Figure 1 - Exemples de polymères synthétiques contenant de l'information binaire. Tous les polymères représentés ici sont des macromolécules uniformes dans lesquelles un alphabet de deux ou plusieurs monomères a été utilisé pour inscrire une séquence binaire. Tous les exemples de cette figure ont été préparés à l'Institut Charles Sadron.

des messages moléculaires qui ressembleraient à ceux que contiennent les chaînes d'ADN. Cette situation a changé depuis environ une dizaine d'années avec l'apparition d'un certain nombre de méthodes permettant d'écrire, de lire et d'éditer les séquences de monomères des copolymères synthétiques [8]. Ce sont ces avancées qui sont brièvement décrites dans les pages qui suivent.

Comment inclure de l'information dans une macromolécule ?

Plusieurs prérequis sont nécessaires pour stocker de l'information sur un polymère. Il faut tout d'abord choisir un langage, et donc un alphabet. Tout est possible. Si l'alphabet génétique est composé de quatre lettres, rien n'empêche de choisir des alphabets plus restreints ou plus larges. Dans le cas du langage binaire, qui est probablement le moyen le plus universel pour stocker différents types de données, l'alphabet est uniquement composé de deux symboles : 0 et 1. Pour transcrire ces symboles à l'échelle moléculaire, il faut donc sélectionner deux monomères ayant des structures moléculaires légèrement différentes. Cela peut être par exemple une différence de masse molaire (ce que nous utilisons le plus souvent), de composition atomique ou de stéréoconfiguration. Il est toutefois important de préciser que le nombre de monomères utilisés pour la synthèse du polymère peut être supérieur au nombre de symboles de l'alphabet choisi. Si deux monomères sont suffisants pour écrire un message binaire, cela peut aussi être réalisé avec un plus grand nombre de monomères (par exemple 4, 8, 16 ou 32).

Une fois que le code moléculaire est sélectionné, il faut trouver une méthode de synthèse permettant d'attacher ces monomères les uns aux autres dans un ordre voulu. Cela signifie que les méthodes de polymérisation conduisant à des séquences aléatoires doivent être proscrites. Il faut donc faire appel à des approches plus contrôlées (mais aussi plus ardues) où les monomères sont assemblés les uns aux autres pas à pas. Ces méthodes dites de polymérisation « multi-étape » ont été décrites en détail dans un précédent article paru dans *L'Actualité Chimique* [9] et ne seront donc pas rediscutées ici. En quelques mots, dans ces approches, les monomères sont attachés un à un, et dans un ordre précis, sur un support solide qui permet de faciliter les purifications après chaque étape. Ces approches sont donc semblables à celles qui sont utilisées

pour la synthèse chimique d'oligopeptides et d'oligonucléotides [6]. Toutefois, une grande partie de notre travail de recherche de ces dernières années a constitué à montrer que les synthèses multi-étapes de polymères synthétiques peuvent être plus aisées que celles des polymères biologiques. En effet, si les synthèses de polymères naturels sont limitées par des contraintes structurales, celles des polymères non naturels peuvent se faire à l'aide d'une palette beaucoup plus large de réactifs et de réactions chimiques. Ainsi, il est possible de sélectionner des procédés de synthèse rapides, efficaces, voire chimiosélectifs.

La figure 1 montre différents exemples de polymères numériques non biologiques qui ont été synthétisés dans notre laboratoire au cours de ces dernières années. La plupart de ces polymères ont été préparés par des approches orthogonales ne requérant pas l'utilisation de chimies de protection. La figure 2 montre par exemple une stratégie de ce type

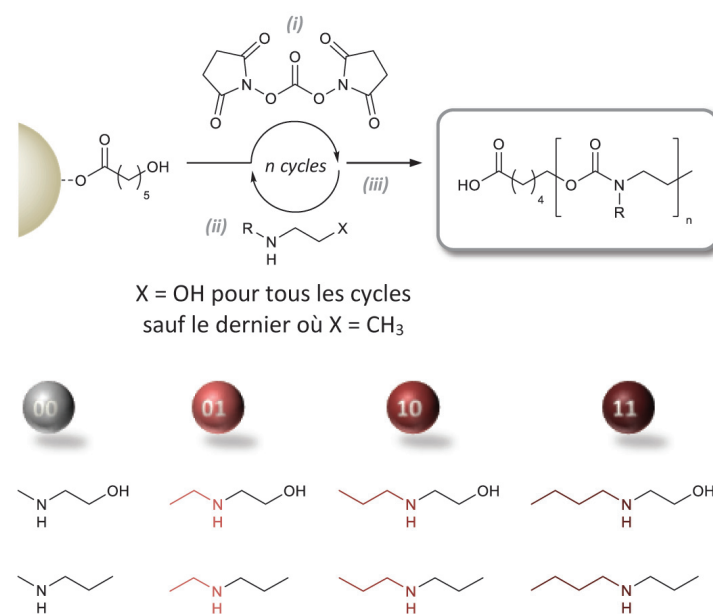


Figure 2 - Synthèse multi-étape d'un polyuréthane N-substitué codé [10]. Ces macromolécules sont obtenues en répétant deux étapes successives : (i) la réaction d'un alcool primaire avec le carbonate de disuccinimidyl, et (ii) la réaction du carbonate activé résultant avec l'amine secondaire d'un aminoalcool. Ces étapes sont répétées n fois jusqu'à obtention d'un polymère de taille souhaitée. Dans le dernier cycle, un agent de terminaison ne contenant pas d'alcool est utilisé dans l'étape (ii). Le codage moléculaire est obtenu en utilisant huit molécules (quatre aminoalcools et quatre amines de terminaison) contenant des substituants alkyles différents.

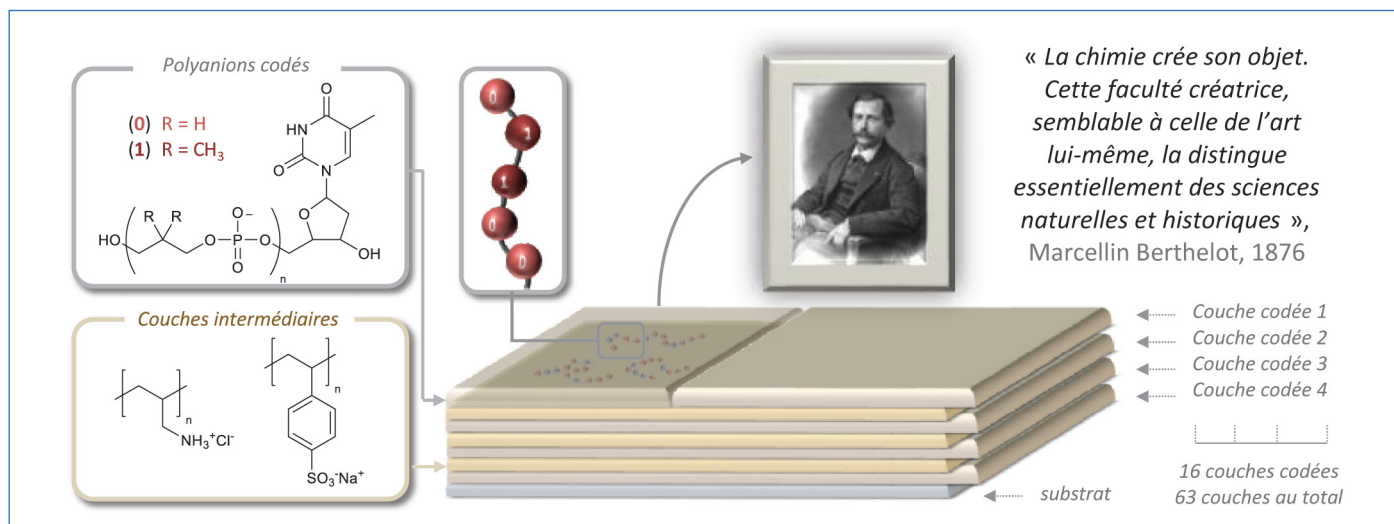


Figure 3 - Stockage 3D d'une chimiothèque de polymères numériques [12]. Dans cet exemple, une citation célèbre de Marcellin Berthelot (1827-1907) a été transcrite en langage binaire et décomposée en fragments qui ont été stockés sur seize poly(phosphodiester)s numériques différents. Ces polymères étant des polyanions, ils ont été ensuite incorporés dans un film multicouche de polyélectrolytes à l'aide de la technique de dépôt couche par couche. Dans cette approche, des strates anioniques et cationiques sont alternées. La déposition successive de ces couches permet de former un film stratifié.

permettant de préparer des polyuréthanes *N*-substitués [10]. Dans cette approche, deux réactions très efficaces (c'est-à-dire à haut rendement) et orthogonales l'une par rapport à l'autre sont répétées successivement jusqu'à obtention d'un polymère uniforme de taille souhaitée. Le codage moléculaire est effectué grâce à quatre monomères différents qui peuvent être employés de manière interchangeable dans une des deux étapes du procédé. Chaque unité codée de la chaîne macromoléculaire formée stocke donc une dyade d'information (c'est-à-dire 00, 01, 10 ou 11).

Dans l'ensemble, les synthèses de polymères codés non biologiques ont été améliorées de manière significative au cours des cinq dernières années. Outre les exemples montrés sur la figure 1, différents autres types de polymères numériques ont été récemment décrits [11]. Il est aussi possible de créer de longues chaînes d'information possédant une centaine d'unités codées [9]. Toutefois, les procédés multi-étapes sont limités par les rendements des réactions successives et ne peuvent pas être répétés indéfiniment. Cependant, même si des chaînes codées géantes ne sont pas envisageables, il est possible d'organiser des chimiothèques de polymères numériques dans l'espace, par exemple sur des surfaces planes, voire même en trois dimensions. La figure 3 montre par exemple l'élaboration d'un film multicouche de polyélectrolytes contenant des strates successives de polymères numériques. Dans ce cas, une chimiothèque de seize polymères codés a été utilisée pour stocker une citation de Marcellin Berthelot [12].

Séquençage de l'information macromoléculaire

Un autre aspect important de ce nouveau domaine de recherche est bien sûr la lecture de l'information macromoléculaire. Parmi les applications possibles des polymères numériques, certaines comme le stockage de données, les technologies anti-contrefaçon ou la cryptographie moléculaire requièrent de pouvoir recouvrer l'information stockée dans un polymère de manière efficace (c'est-à-dire sans erreur de lecture), et si possible en un temps très court. Pour ce faire, des outils analytiques, habituellement utilisés pour séquencer

la structure primaire des biopolymères, ont été revisités pour décoder des polymères synthétiques. Ces approches de séquençage des polymères non naturels ont été présentées très récemment dans une fiche de *L'Actualité Chimique* [13] et ne sont donc pas rediscutées en détail ici. En quelques mots, il existe aujourd'hui peu de méthodes permettant de séquencer efficacement la structure primaire de polymères numériques non naturels. Beaucoup de ces outils sont encore en cours de développement. Par exemple, le séquençage dans des nanopores, qui est une technique établie pour l'analyse de l'ADN, n'a pas encore été complètement adapté à la lecture des polymères synthétiques, même si des recherches intensives sont menées dans ce domaine [13]. À ce jour, la technique la plus efficace pour décoder des polymères synthétiques reste la spectrométrie de masse en tandem (MS/MS). En collaboration avec l'équipe de Laurence Charles d'Aix-Marseille Université, nous avons montré que cette méthode permet de séquencer efficacement tous les polymères numériques montrés dans la figure 1, même si certaines de ces macromolécules sont plus faciles à lire que d'autres. En outre, l'analyse MS/MS de polymères possédant des structures chimiques variées a permis de concevoir des macromolécules « hautement lisibles », chose qu'il est bien sûr difficile de faire avec des polymères naturels. Ainsi, en croisant différentes familles de polymères montrées dans la figure 1, nous avons récemment préparé de longues macromolécules codées qui peuvent être séquencées dans un spectromètre de masse de routine [14].

Manipuler les séquences d'information

Si le stockage macromoléculaire d'information ne requiert en principe que des outils d'écriture et de lecture, la chimie des polymères permet de concevoir d'autres propriétés. En effet, il est aussi possible de modifier une séquence de monomères à l'aide d'un stimulus extérieur qui peut être chimique ou physique. Par exemple, nous avons montré que l'information stockée sur une macromolécule peut être effacée par simple chauffage [15] ou par irradiation UV [16]. Pour obtenir un effaçage thermique, nous avons développé un polymère codé contenant une fonction alcoxyamine thermolabile dans

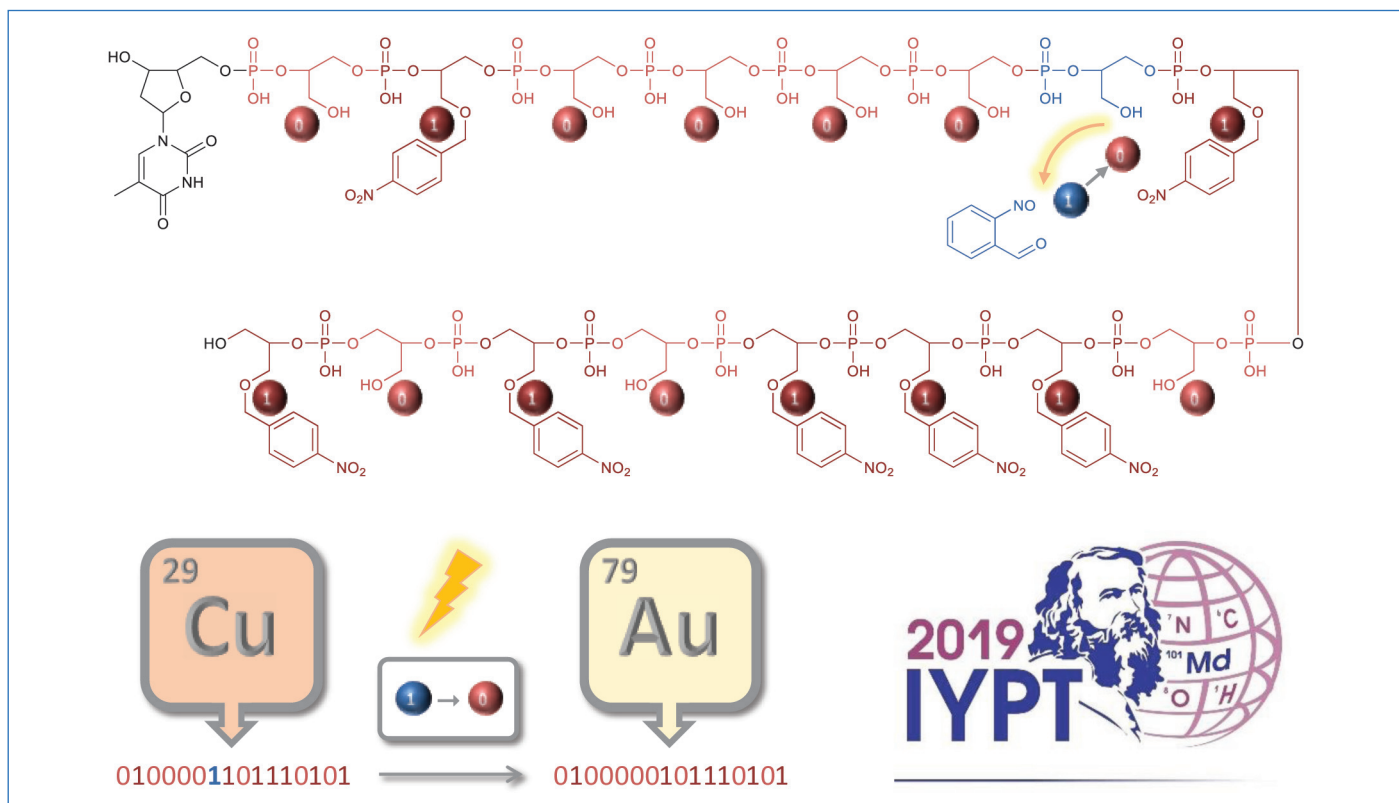


Figure 4 - **Photo-mutation d'un polymère synthétique contenant de l'information binaire** [16]. Le polymère montré en haut de la figure est construit à l'aide de trois monomères. Le monomère contenant une fonction OH (rouge clair) représente un bit-0 d'information. Les monomères contenant des groupes *ortho*-nitrobenzyle (bleu) ou *para*-nitrobenzyle (rouge foncé) sont des isomères de masse et représentent tous deux un bit-1. Lorsque le polymère est irradié, le monomère *ortho*-nitrobenzyle est le seul site photo-modifié qui se transforme en un bit-0. Le polymère dans cet exemple stocke au départ la séquence binaire 0100001101110101, qui est la représentation ASCII du symbole atomique du cuivre, Cu. Après photo-irradiation, la séquence se transforme en 0100000101110101, qui est le codage ASCII du symbole atomique de l'or, Au. Cet exemple de « transmutation numérique » a été conçu pour fêter l'Année internationale du tableau des éléments (IYPT 2019).

chaque unité répétitive. Ce polymère est stable à température ambiante mais se dégrade intégralement lorsqu'il est chauffé au-dessus de 60 °C [15]. Dans le cas d'un effaçage par irradiation, les polymères numériques contiennent des unités codées photolabiles, de type *ortho*-nitrobenzyle (*o*-NB) et *ortho*-nitroveratryle, qui sont toutes clivées lorsqu'elles sont exposées à la lumière.

Mais l'effaçage d'une séquence d'information macromoléculaire n'est qu'une facette de ce qui peut être réalisé. Des modifications plus subtiles peuvent être imaginées. Par exemple, nous avons récemment montré que l'irradiation UV peut aussi être utilisée pour révéler une séquence d'information [16]. Dans ce cas, deux monomères de même masse molaire, contenant soit un motif *o*-NB soit un motif *para*-nitrobenzyle (*p*-NB), ont été utilisés comme langage moléculaire. Cet alphabet ne peut pas être décrypté par MS/MS car les deux monomères sont des isomères de masse. Toutefois, lorsque le polymère est irradié, seuls les motifs *o*-NB sont altérés. La masse molaire de ces unités devient donc plus faible que celles des unités *p*-NB et le polymère devient ainsi lisible. Ce langage moléculaire est donc une « encre sympathique » permettant de cacher un message dans une macromolécule.

Il est même possible d'aller plus loin et d'induire des mutations ciblées dans des séquences d'information. La figure 4 montre un exemple amusant de « transmutation numérique » où une séquence binaire représentant Cu, le symbole chimique du cuivre, a été transformée en celle représentant Au, le symbole de l'or. Cette expérience a été conçue pour célébrer l'Année internationale du tableau périodique des éléments chimiques (IYPT 2019). Pour ce faire, trois

monomères différents ont été utilisés pour construire un poly(phosphodiester) numérique photosensible. L'un d'eux, contenant une fonction alcool, représente un 0, alors que les deux monomères sus-cités, contenant les isomères *o*-NB et *p*-NB, représentent tous deux un 1. Sous exposition UV, le monomère *o*-NB est le seul à être modifié et il se transforme alors en unité 0. Dans l'exemple montré sur la figure 4, cette mutation ciblée 1 → 0 a été volontairement implémentée sur la septième position codée de la séquence (en partant de l'extrémité thymine) afin d'obtenir la modification souhaitée du message binaire.

Au-delà de cet exemple d'alchimie numérique, ce type de modification représente une réelle avancée dans le domaine des polymères synthétiques. En effet, cette chimie est inspirée par les phénomènes de photo-mutagenèse de l'ADN et pourrait être un premier pas vers le développement de polymères synthétiques adaptatifs, c'est-à-dire vers des macromolécules pouvant adapter leur structure et leurs propriétés à des changements de conditions [17]. Pour aller même plus loin, il serait envisageable d'induire des phénomènes évolutifs dans des polymères non naturels. Cette possibilité est discutée plus avant ci-après.

Copier, retranscrire : vers une biologie artificielle ?

L'écriture, la lecture et la modification de séquences macromoléculaires d'information sont donc devenues des réalités scientifiques au cours de ces quelques dernières années. En plus de leur intérêt fondamental, ces nouveaux travaux ont aussi un réel potentiel industriel, notamment dans le domaine

de la lutte anti-contrefaçon où les polymères codés peuvent être utilisés comme des étiquettes moléculaires dispersées à l'état de trace dans des matériaux hôtes. Ces recherches appliquées sont désormais très avancées mais ne peuvent pas être discutées en détail ici pour des raisons de confidentialité. Au-delà des progrès fondamentaux et applicatifs qui ont déjà été effectués avec les polymères numériques (et plus généralement avec les polymères codés), le contrôle de l'information macromoléculaire pourrait aussi ouvrir de nouveaux horizons pour le développement de biologies artificielles rudimentaires [18]. Si les propriétés de l'ADN essentielles à la vie n'ont à ce jour jamais été recréées sur des polymères synthétiques, mis à part sur des xéno-acides nucléiques dont la structure moléculaire est en général très proche de celle des polymères génétiques naturels, il est plus que jamais permis d'en rêver. Pour atteindre un tel but, il faudrait concevoir des macromolécules synthétiques combinant au moins les trois propriétés suivantes : stockage d'information, transfert d'information, mutabilité. Les résultats présentés dans cet article montrent que l'élaboration de tels polymères ne relève plus de la science-fiction. En effet, le stockage d'information macromoléculaire peut désormais se faire très facilement et des exemples sommaires de mutations synthétiques ont été décrits plus haut. Par contre, le transfert d'information d'une macromolécule synthétique à une autre reste un challenge important, même s'il existe déjà de nombreux exemples de petites molécules non biologiques pouvant se dupliquer [18]. Ainsi, les polymères contenant de l'information pourraient être un premier pas vers la création de la vie artificielle, c'est-à-dire vers une abiogénèse induite. Nos travaux de recherche s'orientent déjà dans ce sens.

- [1] Hilbert M., López P., The world's technological capacity to store, communicate, and compute information, *Science*, **2011**, 332, p. 60.
 [2] Lander E.S. et al., Initial sequencing and analysis of the human genome, *Nature*, **2001**, 409, p. 860.
 [3] Venter J.C. et al., The sequence of the human genome, *Science*, **2001**, 291, p. 1304.
 [4] Gibson D.G. et al., Creation of a bacterial cell controlled by a chemically synthesized genome, *Science*, **2010**, 329, p. 52.

- [5] Jinek M. et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science*, **2012**, 337, p. 816.
 [6] Lutz J.-F., Ouchi M., Liu D.R., Sawamoto M., Sequence-controlled polymers, *Science*, **2013**, 341, p. 1238149.
 [7] Dawkins R., *The Blind Watchmaker*, W.W. Norton & Company Inc., New York, **1986**; trad. française : Robert Laffont, **1989**.
 [8] Lutz J.-F., Plastiques binaires : l'avenir de la mémoire numérique ?, *La Recherche*, **2017**, 523, p. 60.
 [9] Lutz J.-F., Les polymères codés : une nouvelle propriété de la matière synthétique, *L'Act. Chim.*, **2016**, 404, p. 16.
 [10] Mondal T., Greff V., Petit B.E., Charles L., Lutz J.-F., Efficient protocol for the synthesis of "N-coded" oligo- and poly(N-substituted urethanes), *ACS Macro Lett.*, **2019**, 8, p. 1002.
 [11] Rutten M.G.T.A., Vaandrager F.W., Elemans J.A.A.W., Nolte R.J.M., Encoding information into polymers, *Nat. Rev. Chem.*, **2018**, 2, p. 365.
 [12] Szweda R., Tschopp M., Felix O., Decher G., Lutz J.-F., Sequences of sequences: spatial organization of coded matter through layer-by-layer assembly of digital polymers, *Angew. Chem. Int. Ed.*, **2018**, 57, p. 15817.
 [13] Charles L., Lutz J.-F., Fiche n° 56 : Le séquençage des polymères numériques, *L'Act. Chim.*, **2018**, 429, p. 63, téléchargement libre : www.lactualitechimique.org/Fiche-no-56-Le-sequençage-des-polymeres-numeriques
 [14] Al Ouahabi A., Amalian J.A., Charles L., Lutz J.-F., Mass spectrometry sequencing of long digital polymers facilitated by programmed inter-byte fragmentation, *Nat. Commun.*, **2017**, 8.
 [15] Roy R.K. et al., Design and synthesis of digitally encoded polymers that can be decoded and erased, *Nat. Commun.*, **2015**, 6, p. 7237.
 [16] König N.F. et al., Photo-editable macromolecular information, *Nat. Commun.*, **2019**, 10, p. 3774.
 [17] Lehn J.-M., Perspectives in chemistry: aspects of adaptive chemistry and materials, *Angew. Chem. Int. Ed.*, **2015**, 54, p. 3276.
 [18] Lutz J.-F., Can life emerge from synthetic polymers?, soumis à *Isr. J. Chem.*

Jean-François LUTZ*

Directeur de recherche au CNRS, Université de Strasbourg, CNRS UPR 22, Institut Charles Sadron, Strasbourg.

Il a reçu la Médaille d'argent du CNRS en 2018.

*jflutz@unistra.fr

© Nicolas BUSSIER/CNRS Photothèque.



ESAT 2020 - 31^e congrès européen sur la thermodynamique
28 juin – 1er juillet 2020 ■ Paris - France

Appel à communications en cours

Cette nouvelle édition mettra à l'honneur

- les dernières avancées expérimentales,
- la modélisation des propriétés thermophysiques des systèmes complexes,
- la thermodynamique moléculaire et statistique,
- l'apport des outils de chemo-informatiques pour la modélisation.



Une opportunité unique de **partager, échanger, débattre** sur les tous derniers développements et de créer de **nouvelles interactions et collaborations scientifiques**.

Plénières, sessions parallèles, sessions posters, table ronde, stands industriels, Job forum.

Co-organisé par



www.esat2020.com
ESAT2020@ifpen.fr

